The
PSYCHOLOGICAL RECORD

Vol. I AUGUST, 1937 No. 14

DIFFERENTIAL RESPONSES TO PERSONALITY TEST ITEMS

MASON N. CROOK

The Principia Press
Bloomington, Ind.

Price of this number, 15 cents

The Principia Press, Inc., has undertaken the publication of this co-operative journal to afford authors the opportunity of immediate publication at the least possible expense. The present low cost of publication, and possible future reductions, depend entirely upon the number of subscriptions. The subscription price has been made low to induce individuals to subscribe. Under the Articles of Incorporation of the Principia Press no profit can be made on any of its publications. Therefore an increase in the number of subscribers will be reflected in reduced costs to authors and an increase in the number of pages published annually.

EDITORIAL BOARD. The above named board of associate editors have agreed to serve in an active capacity for a period of three years, and all manuscripts will be submitted to them.

MANUSCRIPTS may be sent to the appropriate associate editor or to Dr. J. R. Kantor. Longer papers (30 or more typewritten pages) in the above mentioned fields are especially desired.

COSTS. The cost of publication will be $2.00 per page of approximately 500 words. For tabular matter or illustrations in excess of 10 per cent of the total length of the paper there will be an excess charge of 75 cents per page. An additional 50 cents will be charged for each paper for the cost of covers.

REPRINTS. One hundred copies of each paper will be supplied gratis. Additional copies may be purchased in any quantity at a rate which will be set when the author is informed of the cost of publication.

ROYALTIES. Fifty per cent of the net income from the sale of individual copies, or from copies sold as part of back volumes, will be credited as royalties to the author's account. Royalties cannot be paid on income from subscriptions current in the year of publication.

SUBSCRIPTIONS. The subscription price is $4.00 per year for an annual volume of approximately 500 pages. Individual papers will be sold at an advertised price, which will depend upon the size of the article. Foreign subscriptions will be $4.50.

CORRESPONDENCE concerning editorial matters should be addressed to Dr. J. R. Kantor, Indiana University. Business correspondence should be addressed to Dr. C. M. Louttit, Principia Press, Inc., Bloomington, Ind.

DIFFERENTIAL RESPONSES TO PERSONALITY TEST ITEMS*

MASON N. CROOK
Skidmore College

The typical questionnaire type of personality test consists of numerous items, all of which contribute more or less to the total score in the trait which the test is supposed to measure. It is basic to the theory of test construction, however, that such items shall pertain to different specific situations. As a consequence, it is found that items differ considerably among themselves in such statistical properties as incidence in the population and diagnostic significance. The purpose of this note is twofold: (1) to present graphically the distribution of responses among the various response categories on each item of a personality test, as such distributions afford a particularly effective demonstration of the difference between items, and (2) to point out certain considerations bearing on the nature of these distributions.

In a test calling for "yes" or "no" responses, the items of high incidence will show a larger percentage of "yes" responses than the others. In a test calling for marking on a continuous scale, or offering a graduated series of choices, a more refined picture of this difference between items can be obtained. The data here reported are from a test of this latter type. The Willoughby (Clark-Thurstone) Personality Schedule is a test for neuroticism consisting of 25 items, on each of which the subject has a choice of 5 responses. The items are as follows:

1. Do you get stage fright? 2. Do you worry over humiliating experiences? 3. Are you afraid of falling when you are on a high place? 4. Are your feelings easily hurt? 5. Do you keep in the background on social occasions? 6. Are you happy and sad by turns without knowing why? 7. Are you shy? 8. Do you daydream frequently? 9. Do you get discouraged easily? 10. Do you say things on the spur of the moment and then regret them? 11. Do you like to be alone? 12. Do you cry easily? 13. Does it bother you to have people watch you work even when you do it well? 14. Does criticism hurt you badly? 15. Do you cross the street to avoid meeting someone? 16. At a reception or tea do

* Manuscript recommended for publication by Dr. J. R. Kantor, July 3, 1937.

you avoid meeting the important person? 17. Do you often feel just miserable? 18. Do you hesitate to volunteer in a class recitation? 19. Are you often lonely? 20. Are you self-conscious before superiors? 21. Do you lack self-confidence? 22. Are you self-conscious about your appearance? 23. If you see an accident does something keep you from giving help? 24. Do you feel inferior? 25. Is it hard for you to make up your mind until the time for action is past?

The subject indicates his response by encircling one of the numbers from 0 to 4, which in the regular blank are printed after each item. The key is as follows:

0 means "no", "never", "not at all", etc.; 1 means "somewhat", "sometimes", "a little", etc.; 2 means "about as often as not", "an average amount", etc.; 3 means "usually", "a good deal", "rather often", etc.; 4 means "practically always", "entirely", etc.

The test was given to 226 men and 321 women students, mostly sophomores, at the University of California at Los Angeles. Items were read aloud by the experimenter and responses entered on record sheets prepared in advance, with a time limit of 15 seconds on each item. Each subject recorded name of maternal grandparent of same sex, rather than own name, to give a degree of anonymity. Frequencies for the various response categories, on each of the 25 items, are presented in Fig. 1 in terms of percentages. The heavy solid line represents men and the heavy dashed line women. The faint lines will be discussed later.

It should be borne in mind that the response categories cannot be assumed to give a scale of equal units on the base line. The line graphs are used instead of column diagrams for the sake of clarity and compactness, but they do not represent functional relationships between continuous variables in the strict mathematical sense. In this connection, it may be pointed out that various problems of scaling are related to the data here presented, but a consideration of them is beyond the scope of this discussion.

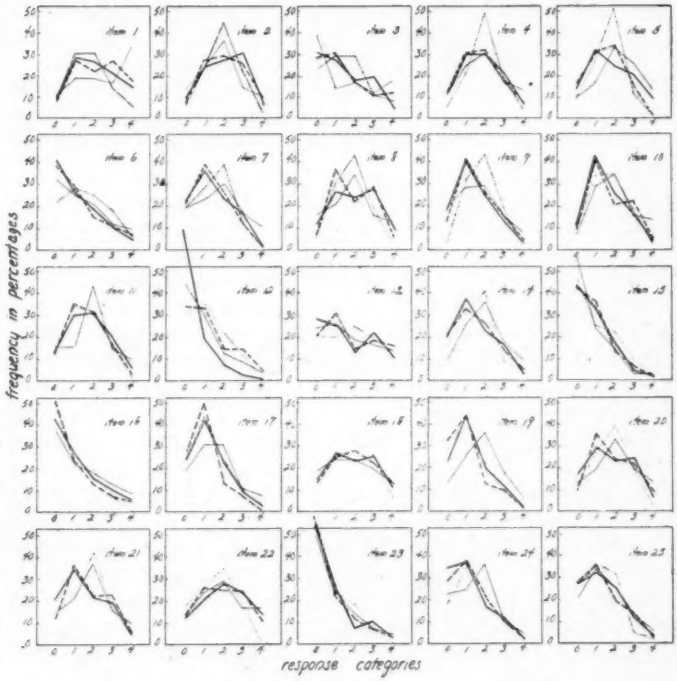
Certain features of the distributions represented by the heavy lines in Fig. 1 deserve special mention:

1. The modal response varies from category 0 on some items to category 2 on others. On the latter type of item the distributions appear approximately symmetrical. It is with this difference between items that we are mainly concerned.

2. There is a remarkably close agreement between men and women, even to a certain amount of bimodality on items 8, 13, 20,

FIGURE 1

Distributions of responses on the 25 items of the Willoughby Personality Schedule. ——— California male Students; — — — California female students; ..... Skidmore female students; ——— unmarried professional women (after Willoughby & Morse).



and 21. Largest sex difference appears on item 12, "Do you cry easily?". This comparison between the sexes will be discussed later.

On the items showing a modal response at 0 or 1 the majority of people give themselves a more favorable rating than the mid-point on the scale. Does this type of response mean that the majority of subjects believe themselves to possess the traits in question to a lesser extent than the average person? Such an interpretation would be consistent with much other evidence that compensatory dynamisms influence questionnaire responses. (1, 5). The formal instructions, however, leave some uncertainty of interpretation. A score of

2 means "about as often as not", "an average amount", etc. A subject might put the emphasis on "an average amount" and assume it to pertain to the average of the population. On the other hand, he might put the emphasis on "about as often as not" and assume that the average person would manifest the behavior in question less often than "about as often as not". If the majority of individuals in a population adopted the former interpretation a modal response of 0 or 1 by that population would signify a systematic tendency on the part of the subjects to rate themselves more favorably relative to the average than they deserved. If a majority of the individuals adopted the latter interpretation a modal response of 0 or 1 would signify no such tendency.

To clarify this point the test was given, with instruction so modified as to force the former interpretation, to 127 women students, mostly sophomores, at Skidmore College. Administration differed slightly from than in the California testing. The blanks were passed out in class and the subjects were told to read the instructions and to enter the date and mother's maiden name, but not to mark the items until the signal was given. The following supplementary instructions were then read verbatim:

> "A little explanation of the instructions is necessary before we start to mark the blanks. Each item is to be marked by encircling one of the numbers—0, 1, 2, 3, 4. The number 2 is considered average—that is, 2 is the number that would be encircled by the average person, and therefore the number that would be encircled most often. Therefore you mark the 2 on an item if you think you are about an average person with respect to that item; mark 0 or 4 if you differ very much from the average in one direction or the other, and mark 1 or 3 for intermediate degrees."

The signal to proceed was given and subjects marked the items with no time limit.

Results of this testing are plotted in the light dotted lines. The noteworthy features of this set of curves can be stated in 2 main points:

1. There is a tendency for the mode to settle on response 2, which in some cases serves merely to sharpen the mode (items 2, 4, 5, 22), in other cases serves to eliminate bimodality (items 8, 20, 21), and in still others involves a clear shift of the mode from a lower category (items 7, 9, 14). These changes can probably be attributed to the special instruction, though the difference between the populations should not be entirely disregarded.

2. In spite of the type of shift described above, items which in the original population had the greatest concentration of responses on the low categories show relatively little change. In terms of the original population, the 5 most extreme items in this respect selected by inspection are 6, 12, 15, 16, and 23. We shall refer to these items as group A. In the new population the same five items hold the same extreme position, and the last 3 of the 5 show practically no change. This result is sufficiently unambiguous to justify the conclusion that on certain items there is a systematic tendency for college students to give themselves more favorable ratings relative to the average than they deserve.

An explanation is suggested by the content of the items. The 5 items in group A are: 6 (mood swings), 12 (cry easily), 15 (cross street to avoid someone), 16 (avoid important person), and 23 (revulsion from accident). With these we can contrast the 5 items which appear to show, relative to other items, the greatest concentration of responses on the higher categories: 1 (stage fright), 2 (humiliation), 4 (feelings hurt), 18 (hesitate in recitation), and 22 (self-conscious about appearance). These we shall call group B.

The difference between the two groups seems to lie, not in the general type of situation, but rather in the connotation of undesirability as judged by the naive subject. Under this interpretation, all the items in group A presumably imply either mental instability, as No. 6, or a mild social stigma, as No. 23, and in consequence positive responses are inhibited. In group B, on the other hand, the items would seem to imply either a mere personal foible, as No. 18, or behavior justified on the ground of unfamilarity with the situation, as No. 1, or even commendable conformity with social usage, as No. 22.

If "yes" or "no" responses were required, items of group B should, of course, show higher incidence in the population than items of group A. All items in the Willoughby test appear in "yes" or "no" form in the original Thurstone test. The 41 Thurstone items of highest incidence in a Texas student population have been listed by Harvey (2), and the 40 items of highest incidence in a population of Pittsburgh freshman women have been listed by Willoughby (3). Of our B group, all items appear on both the Harvey and Willoughby lists, while of our A group only 3 appear on the Harvey list and 4 on the Willoughby. This difference, while not striking, is in the expected direction. If the interpretation of the item differences offered above is correct, then the obtained

incidence score for an item depends perhaps as much on the degree of insight of the subjects as on the actual facts of behavior.

It was pointed out previously that the men and women in our California population agree very closely. Data are reported by Willoughby (4) showing that the average response on an individual item varies widely as a function of age and marital status. These data further suggest that sex differences are in general greater at later ages than during the college years, and therefore the close similarity here found between the sexes is no doubt in part a function of the population.

It also appears from the Willoughby data, however, that the nature of the variation with age and marital status depends on the particular item. In view of these complicating considerations, it would be of interest to know whether the above analysis of item differences in a student population would be borne out in a population of different type. In other words, it would be desirable to know to what extent the relation of the items to each other, in respect to distribution of responses among the choice categories, is a function of the population; our knowledge that the average response on an item is a function of the population as well as of the item content does not permit a conclusive inference on this point. Results which afford at least a partial answer to the question have been reported by Willoughby and Morse (5), on a population of about 80 unmarried professional women, with a small scattering of bachelors and other special types. Ages were concentrated mostly in the range from the late 40's to the early 60's. These data are represented in the graph by the light solid lines. On some items (6, 8, 10, 14, 15, 16, 19, 23) the wording used by Willoughby and Morse was slightly, but probably insignificantly, different from that on the published form used in the present study. On items 1 and 18 there was perhaps sufficient difference to deserve separate mention. Item 1 appears in the published version, "Do you get stage fright?", in the Willoughby and Morse version, "Does it make you nervous to have to talk to an audience?"; item 18, published version, "Do you hesitate to volunteer in a class recitation?", Willoughby and Morse version, "Do you hesitate to express yourself in a group discussion?" All other items are identical in wording in the two lists.

Fairly large differences between these curves and those from the student populations are apparent on certain items, as was to be expected from the evidence on the effect of age discussed above. It is interesting that these results, which were obtained by individual

interview, are somewhat more similar to our Skidmore data, obtained under the special instruction, than to our California data.

The most significant analysis, however, is with respect to our extreme groups of items. On inspection of the curves we find that the items of group A (6, 12, 15, 16, 23) show the greatest concentration of responses on the low categories for the unmarried professional women as they do for the students, and the items of group B (1, 2, 4, 18, 22), while not constituting quite so unambiguously the other extreme, at least conform fairly well, in the older population, to their general pattern in the student group. Item No. 1 shows a shift of mode to category 4, which is consistent with the age trend found by Willoughby (4) for this item, but which may also be due in part to the large change in wording, for reasons suggested in the last paragraph below.

Thus, in spite of the fact that average level of response on a given item varies widely with age and sex, we have evidence that items differ from each other in respect to distribution of responses among the various choice categories in a way that is to a fair extent independent of age and sex as well as of method of test administration.

From the standpoint of scaling and weighting, probably a selection of items statistically more homogeneous than most tests contain would be advantageous. From the standpoint of clinical value the implication is not so clear. Items which not only present different situations, but call into play different dynamisms in the subjects, would probably afford a broader basis for analysis and diagnosis. On the other hand, a completely disguised test on the order of the Watson Test of Public Opinion (fairmindedness) might provide more valid total scores, and the interpretation here offered suggests that the high incidence items afford a possible technique to this end.

A concluding point of some interest is the large significance of what might appear superficially to be minor differences in situations; e. g., item 16, "At a reception or tea do you avoid meeting the important person?" and item 5, "Do you keep in the background on social occasions?". Item 16 is in group A; item 5, while not formally in group B, shows a response pattern very similar to the items of group B, and very different from those of group A. It would obviously be difficult to make a reliable a priori judgment what kind of difference between the responses on these two items to expect, and only a theory based on a larger number of more easily analyzable items would permit us to infer with any assurance that item 16

suggests to the layman an active recession from a social situation, while item 5 implies merely a becoming degree of modesty.

SUMMARY

Distributions of responses among the 5 response categories on the Willoughby Personality Schedule, obtained from several populations, are graphically presented for each of the 25 items. It is pointed out that the distribution pattern characterizing a given item is to a certain extent independent of age, sex, and method of test administration, and evidence is offered that on certain items the shape of the distribution is determined largely by a systematic tendency for the subjects to give themselves more favorable ratings than they deserve.

REFERENCES

1. Crook, M. N. The constancy of neuroticism scores and self-judgments of constancy. *J. Psychol.*, 1937, 4, 27-34.
2. Harvey, O. L. Concerning the Thurstone "Personality Schedule". *J. Soc. Psychol.*, 1932, 3, 240-251.
3. Willoughby, R. R. Some properties of the Thurstone Personality Schedule and a suggested revision. *J. Soc. Psychol.*, 1932, 3, 401-424.
4. Willoughby, R. R. Emotionality, age, sex, and conjugal condition. *Amer. J. Sociol.*, in press.
5. Willoughby, R. R., & Morse, M. E. Spontaneous reactions to a personality inventory. *Amer. J. Orthopsychiat.*, 1936, 6, 562-575.